



- (51) **International Patent Classification:**
G06F 13/00 (2006.01)
- (21) **International Application Number:**
PCT/US201 1/067845
- (22) **International Filing Date:**
29 December 2011 (29.12.2011)
- (25) **Filing Language:** English
- (26) **Publication Language:** English
- (71) **Applicant (for all designated States except US):** INTEL CORPORATION [US/US]; 2200 Mission College Blvd., P.O.Box 581 19, Santa Clara, CA 95052-81 19 (US).
- (72) **Inventors; and**
- (75) **Inventors/Applicants (for US only):** STEELY, Simon, C. [US/US]; 8 Anna Louise Dr., Hudson, NH 03051 (US). HASENPLAUGH, William, C. [US/US]; 44 St. Botolph St. Apt. ID, Boston, MA 02116 (US). EMER, Joel, S. [US/US]; 77 Reed Road, Hudson, MA 01749 (US).
- (74) **Agents:** O'DOWD, Shawn et al; Kenyon & Kenyon LLP, 1500 K Street, N.W., Suite 700, Washington, DC 20005 (US).
- (81) **Designated States (unless otherwise indicated, for every kind of national protection available):** AE, AG, AL, AM,

AO, AT, AU, AZ, BA, BB, BG, BH, BR, BW, BY, BZ, CA, CH, CL, CN, CO, CR, CU, CZ, DE, DK, DM, DO, DZ, EC, EE, EG, ES, FI, GB, GD, GE, GH, GM, GT, HN, HR, HU, ID, IL, IN, IS, JP, KE, KG, KM, KN, KP, KR, KZ, LA, LC, LK, LR, LS, LT, LU, LY, MA, MD, ME, MG, MK, MN, MW, MX, MY, MZ, NA, NG, NI, NO, NZ, OM, PE, PG, PH, PL, PT, QA, RO, RS, RU, RW, SC, SD, SE, SG, SK, SL, SM, ST, SV, SY, TH, TJ, TM, TN, TR, TT, TZ, UA, UG, US, UZ, VC, VN, ZA, ZM, ZW.

- (84) **Designated States (unless otherwise indicated, for every kind of regional protection available):** ARIPO (BW, GH, GM, KE, LR, LS, MW, MZ, NA, RW, SD, SL, SZ, TZ, UG, ZM, ZW), Eurasian (AM, AZ, BY, KG, KZ, MD, RU, TJ, TM), European (AL, AT, BE, BG, CH, CY, CZ, DE, DK, EE, ES, FI, FR, GB, GR, HR, HU, IE, IS, IT, LT, LU, LV, MC, MK, MT, NL, NO, PL, PT, RO, RS, SE, SI, SK, SM, TR), OAPI (BF, BJ, CF, CG, CI, CM, GA, GN, GQ, GW, ML, MR, NE, SN, TD, TG).

Declarations under Rule 4.17:

— *of inventorship (Rule 4.17(iv))*

Published:

— *with international search report (Art. 21(3))*

(54) **Title:** DOMAIN STATE

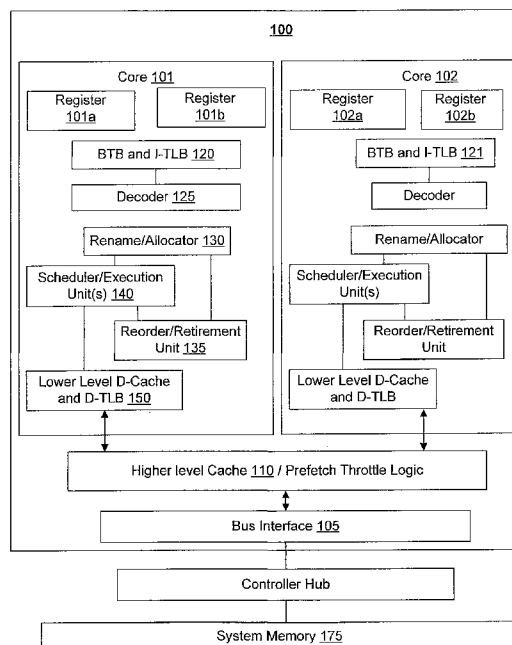


FIG. 1

(57) **Abstract:** Method and apparatus to efficiently maintain cache coherency by reading/writing a domain state field associated with a tag entry within a cache tag directory. A value may be assigned to a domain state field of a tag entry in a cache tag directory. The cache tag directory may belong to a hierarchy of cache tag directories. Each tag entry may be associated with a cache line from a cache belonging to a first domain. The first domain may contain multiple caches. The value of the domain state field may indicate whether its associated cache line can be read or changed.

DOMAIN STATE

FIELD OF THE INVENTION

5 **[0001]** The present disclosure pertains to the field of processors and, in particular, to optimizing cache management techniques.

DESCRIPTION OF RELATED ART

10 **[0002]** Advances in semi-conductor processing and logic design have permitted an increase in the amount of logic that may be present on integrated circuit devices. As a result, computer system configurations have evolved from a single or multiple integrated circuits in a system to multiple cores, multiple hardware threads, and multiple logical processors present on individual integrated circuits. A processor or integrated circuit typically comprises a single physical processor die, where the
15 processor die may include any number of cores, hardware threads, or logical processors. The ever increasing number of processing elements—cores, hardware threads, and logical processors—on integrated circuits enables more tasks to be accomplished in parallel. However, the execution of more threads and tasks put an increased premium on shared resources, such as memory, and the management
20 thereof.

25 **[0003]** Typically, cache memory includes a memory between a shared system memory and execution units of a processor to hold information in a closer proximity to the execution units. In addition, cache is typically smaller in size than a main system memory, which allows for the cache to be constructed from expensive, faster memory, such as Static Random Access Memory (SRAM). Both the proximity to the
30 execution units and the speed allow for caches to provide faster access to data and instructions. Caches are often identified based on their proximity from execution units of a processor. For example, a first-level (L1) cache may be close to execution units residing on the same physical processor. Due to the proximity and placement, first level cache is often the smallest and quickest cache. A computer system may also hold

higher-level or further out caches, such as a second level (L2) cache, which may also reside on the processor but be placed between the first level cache and main memory. And a third level (L3) cache may be placed on the processor or elsewhere in the computer system, such as at a controller hub, between the second level cache and
5 main memory.

[0004] Cache coherence is a useful mechanism in multiple processor systems to provide a shared memory abstraction to the programmer. When multiple processors cache a given shared memory location, a cache coherence problem may arise because a copy of the same memory location exists in multiple caches. A cache coherence
10 protocol guarantees that a given memory location has a consistent view across all processors. There are many models of what a consistent view is, and one example is sequential consistency. Another is weak ordering. In each of these models, the coherence protocol prescribes a legal order in which memory locations can be acquired by and updated by various processors.

[0005] One conventional technique for maintaining cache coherency, particularly in distributed systems, is a directory-based cache coherency scheme. Directory-based coherency schemes utilize a centralized tag directory to record the location and the status of cache lines as they exist throughout the system. For example, the tag
15 directory records which processor caches have a copy of the data, and further records if any of the caches have an updated copy of the data. When a processor makes a cache request to the main memory for a data item, the tag directory is consulted to determine where the most recent copy of the data resides. Based on this information, the most recent copy of the cache line is retrieved so that it may be provided to the requesting processor cache memory. The tag directory is then updated to reflect the
20 new status for that cache line. Thus, each cache line read by a processor is accompanied by a tag directory update (i.e., a write). The directory-based cache coherency scheme may include multiple tag directories, and the tag directories may be arranged in a hierarchy. Furthermore, a hierarchical tag directory structure may include any number of levels.

DESCRIPTION OF THE FIGURES

[0006] Embodiments are illustrated by way of example and not limitation in the Figures of the accompanying drawings:

[0007] FIG. 1 illustrates an embodiment of a processor including multiple processing elements.

[0008] FIG. 2 illustrates an embodiment of on-core memory interface logic.

[0009] FIG. 3 illustrates an embodiment of multiple processors and tag directories in a hierarchy.

[0010] FIG. 4 illustrates an embodiment of multiple processors, tag directories in a hierarchy, and a domain state field.

[0011] FIG. 5 illustrates, in an embodiment, various conditions which set the values of the domain state field.

[0012] FIG. 6 illustrates, in an embodiment, various conditions which set the values of the domain state field.

DETAILED DESCRIPTION

[0013] In the following description, numerous specific details are set forth such as examples of specific hardware structures for determining cache lines, reading/writing to domain state fields, determining target caches, and determining domains, as well as placement of such hardware structures, such as at memory ports or at independent cache slices; specific processor units/logic, specific examples of processing elements, etc. in order to provide a thorough understanding of the present invention. It will be apparent, however, to one skilled in the art that these specific details need not be employed to practice the present invention. In other instances, well known components or methods, such as specific counter circuits, alternative multi-core and multi-threaded processor architectures, specific un-core logic, specific memory controller logic, specific cache implementations, specific cache coherency protocols, and specific operational details of microprocessors, have not been described in detail in order to avoid unnecessarily obscuring the present invention.

[0014] Embodiments may be discussed herein which efficiently maintain cache coherency. In particular, embodiments of the present invention pertain to a feature for

reading/writing a domain state field associated with a tag entry within a cache tag directory. In an embodiment, a value may be assigned to a domain state field of a tag entry in a cache tag directory. The cache tag directory may belong to a hierarchy of cache tag directories. Each tag entry may be associated with a cache line from a cache
5 belonging to a domain. The domain may contain multiple caches. The value of the domain state field may indicate whether its associated cache line can be read or changed.

[0015] In an embodiment, the domain state field may be assigned a value indicating that the domain state field's associated domain has the only copies of its associated
10 cache line. In an embodiment, the domain state field may be assigned a value indicating that the domain state field's associated domain is a last accessor of its associated cache line, and at least one other domain has or had a copy of that cache line. In an embodiment, the domain state field may be assigned a value indicating that the domain state field's associated domain is not the last accessor of the cache line, and all
15 domains in the hierarchy of cache tag directories are up to date with memory. In an embodiment, the domain state field may be assigned a value indicating that the domain state field's associated domain is not the last accessor of the cache line, and a global last accessor domain has a copy of the cache line which is not up to date with memory.

[0016] In an embodiment, the domain state field may be assigned a value indicating
20 that all read requests and all write requests for its associated cache line can only be serviced in the domain state field's associated domain. In an embodiment, the domain state field may be assigned a value indicating that all read requests for the cache line can be serviced in the domain state field's associated domain, that no write requests for the cache line can be serviced in the domain state field's associated domain, and that
25 write requests for the cache line must go to a next level of the hierarchy of cache tag directories. In an embodiment, the domain state field may be assigned a value indicating that no read requests or write requests for the cache line can be serviced in the domain state field's associated domain, and that read requests or write requests for the cache line must go to a next level of the hierarchy of cache tag directories.

[0017] Referring to FIG. 1, an embodiment of a processor including multiple cores
30 is illustrated. Processor 100, in one embodiment, includes one or more caches.

Processor 100 includes any processor, such as a micro-processor, an embedded processor, a digital signal processor (DSP), a network processor, or other device to execute code. Processor 100, as illustrated, includes a plurality of processing elements.

[0018] In one embodiment, a processing element refers to a thread unit, a thread slot, a process unit, a context, a logical processor, a hardware thread, a core, and/or any other element, which is capable of holding a state for a processor, such as an execution state or architectural state. In other words, a processing element, in one embodiment, refers to any hardware capable of being independently associated with code, such as a software thread, operating system, application, or other code. A physical processor typically refers to an integrated circuit, which potentially includes any number of other processing elements, such as cores or hardware threads.

[0019] A core often refers to logic located on an integrated circuit capable of maintaining an independent architectural state wherein each independently maintained architectural state is associated with at least some dedicated execution resources. In contrast to cores, a hardware thread typically refers to any logic located on an integrated circuit capable of maintaining an independent architectural state wherein the independently maintained architectural states share access to execution resources. As can be seen, when certain resources are shared and others are dedicated to an architectural state, the line between the nomenclature of a hardware thread and core overlaps. Yet often, a core and a hardware thread are viewed by an operating system as individual logical processors, where the operating system is able to individually schedule operations on each logical processor.

[0020] Physical processor 100, as illustrated in FIG. 1, includes two cores, core 101 and 102. Here, core hopping may be utilized to alleviate thermal conditions on one part of a processor. However, hopping from core 101 to 102 may potentially create the same thermal conditions on core 102 that existed on core 101, while incurring the cost of a core hop. Therefore, in one embodiment, processor 100 includes any number of cores that may utilize core hopping. Furthermore, power management hardware included in processor 100 may be capable of placing individual units and/or cores into low power states to save power. Here, in one embodiment, processor 100 provides hardware to assist in low power state selection for these individual units and/or cores.

[0021] Although processor 100 may include asymmetric cores, i.e. cores with different configurations, functional units, and/or logic, symmetric cores are illustrated. As a result, core 102, which is illustrated as identical to core 101, will not be discussed in detail to avoid repetitive discussion. In addition, core 101 includes two hardware threads 101a and 101b, while core 102 includes two hardware threads 102a and 102b. Therefore, software entities, such as an operating system, potentially view processor 100 as four separate processors, i.e. four logical processors or processing elements capable of executing four software threads concurrently.

[0022] Here, a first thread is associated with architecture state registers 101a, a second thread is associated with architecture state registers 101b, a third thread is associated with architecture state registers 102a, and a fourth thread is associated with architecture state registers 102b. As illustrated, architecture state registers 101a are replicated in architecture state registers 101b, so individual architecture states/contexts are capable of being stored for logical processor 101a and logical processor 101b. Other smaller resources, such as instruction pointers and renaming logic in rename allocator logic 130 may also be replicated for threads 101a and 101b. Some resources, such as re-order buffers in reorder/retirement unit 135, ILTB 120, load/store buffers, and queues may be shared through partitioning. Other resources, such as general purpose internal registers, page-table base register, low level data-cache and data-TLB 115, execution unit(s) 140, and portions of out-of-order unit 135 are potentially fully shared.

[0023] Processor 100 often includes other resources, which may be fully shared, shared through partitioning, or dedicated by/to processing elements. In FIG. 1, an embodiment of a purely exemplary processor with illustrative logical units/resources of a processor is illustrated. Note that a processor may include, or omit, any of these functional units, as well as include any other known functional units, logic, or firmware not depicted. As illustrated, processor 100 includes a branch target buffer 120 to predict branches to be executed/taken and an instruction-translation buffer (I-TLB) 120 to store address translation entries for instructions.

[0024] Processor 100 further includes decode module 125 is coupled to fetch unit 120 to decode fetched elements. In one embodiment, processor 100 is associated with an Instruction Set Architecture (ISA), which defines/specifies instructions executable

on processor 100. Here, often machine code instructions recognized by the ISA include a portion of the instruction referred to as an opcode, which references/specifies an instruction or operation to be performed.

[0025] In one example, allocator and renamer block 130 includes an allocator to reserve resources, such as register files to store instruction processing results. However, threads 101a and 101b are potentially capable of out-of-order execution, where allocator and renamer block 130 also reserves other resources, such as reorder buffers to track instruction results. Unit 130 may also include a register renamer to rename program/instruction reference registers to other registers internal to processor 100.

Reorder/retirement unit 135 includes components, such as the reorder buffers mentioned above, load buffers, and store buffers, to support out-of-order execution and later in-order retirement of instructions executed out-of-order.

[0026] Scheduler and execution unit(s) block 140, in one embodiment, includes a scheduler unit to schedule instructions/operation on execution units. For example, a floating point instruction is scheduled on a port of an execution unit that has an available floating point execution unit. Register files associated with the execution units are also included to store information instruction processing results. Exemplary execution units include a floating point execution unit, an integer execution unit, a jump execution unit, a load execution unit, a store execution unit, and other known execution units.

[0027] Lower level data cache and data translation buffer (D-TLB) 150 are coupled to execution unit(s) 140. The data cache is to store recently used/operated on elements, such as data operands, which are potentially held in memory coherency states. The D-TLB is to store recent virtual/linear to physical address translations. As a specific example, a processor may include a page table structure to break physical memory into a plurality of virtual pages.

[0028] As depicted, cores 101 and 102 share access to higher-level or further-out cache 110, which is to cache recently fetched elements. Note that higher-level or furtherout refers to cache levels increasing or getting further way from the execution unit(s). In one embodiment, higher-level cache 110 is a last-level data cache—last cache in the memory hierarchy on processor 100—such as a second or third level data

cache. However, higher level cache 110 is not so limited, as it may be associated with or include an instruction cache. A trace cache—a type of instruction cache—instead may be coupled after decoder 125 to store recently decoded traces.

[0029] Note, in the depicted configuration that processor 100 also includes bus interface module 105 to communicate with devices external to processor 100, such as system memory 175, a chipset, a Northbridge, or other integrated circuit. Memory 175 may be dedicated to processor 100 or shared with other devices in a system. Common examples of types of memory 175 include dynamic random access memory (DRAM), static RAM (SRAM), non-volatile memory (NV memory), and other known storage devices.

[0030] FIG. 1 illustrates an abstracted, logical view of an exemplary processor with a representation of different modules, units, and/or logic. However, note that a processor utilizing the methods and apparatus' described herein need not include the illustrated units. And, the processor may omit some or all of the units shown. To illustrate the potential for a different configuration, the discussion now turns to FIG. 2, which depicts an embodiment of processor 200 including an on-processor memory interface module—an un-core module—with a ring configuration to interconnect multiple cores. Processor 200 is illustrated including a physically distributed cache; a ring interconnect; as well as core, cache, and memory controller components. However, this depiction is purely illustrative, as a processor implementing the described methods and apparatus may include any processing elements, style or level of cache, and/or memory, front-side-bus or other interface to communicate with external devices.

[0031] In one embodiment, caching agents 221-224 are each to manage a slice of a physically distributed cache. As an example, each cache component, such as component 221, is to manage a slice of a cache for a collocated core—a core the cache agent is associated with for purpose of managing the distributed slice of the cache. As depicted, cache agents 221-224 are referred to as Cache Slice Interface Logic (CSIL)s; they may also be referred to as cache components, agents, or other known logic, units, or modules for interfacing with a cache or slice thereof. Note that the cache may be any level of cache; yet, for this exemplary embodiment, discussion focuses on a last-level cache (LLC) shared by cores 201-204.

[0032] Much like cache agents handle traffic on ring interconnect 250 and interface with cache slices, core agents/components 211-214 are to handle traffic and interface with cores 201-204, respectively. As depicted, core agents 221-224 are referred to as Processor Core Interface Logic (PCIL)s; they may also be referred to as core components, agents, or other known logic, units, or modules for interfacing with a processing element. Additionally, ring 250 is shown as including Memory Controller Interface Logic (MCIL) 230 and Graphics Hub (GFX) 240 to interface with other modules, such as memory controller (IMC) 231 and a graphics processor (not illustrated). However, ring 250 may include or omit any of the aforementioned modules, as well as include other known processor modules that are not illustrated. Additionally, similar modules may be connected through other known interconnects, such as a point-to-point interconnect or a multi-drop interconnect.

[0033] It's important to note that the methods and apparatus' described herein may be implemented in any cache at any cache level, or at any processor or processor level. Furthermore, caches may be organized in any fashion, such as being a physically or logically, centralized or distributed cache. As a specific example, the cache may include a physical centralized cache with a similarly centralized tag directory, such as higher level cache 110. Alternatively, the tag directories may be either physically and/or logically distributed in a physically distributed cache, such as the cache organization illustrated in FIG. 2.

[0034] In an embodiment, a processor, such as the processor illustrated in FIG. 1, illustrated in FIG. 2, or any other processor, includes one or more caches. FIG. 3 illustrates an embodiment processor(s) with multiple cores/un-cores. Each processor may include at least a core 312 and at least an un-core 310. In an embodiment, 302, 304, 306, and 308 may be cores/un-cores which are part of a single processor. In another embodiment, 302, 304, 306, and 308 maybe multiple processors. The core 312 may include components of the processor involved in executing instructions. The un-core 310 may include all logic not in the core 312, but which are essential for core performance. The core 312 may include components such as an LI instruction cache (LH) 314 and a LI data cache (LID) 316. The un-core 310 may include components such as a missing address file (MAF) 317, victim buffer (VB) 318, and L2 cache (L2)

322. One or more processors, one or more cores, or one or more un-cores, and their caches may be associated with a domain. In an embodiment, illustrated in FIG. 3, processors 302 and 304, and their cores, un-cores, and caches are associated with domain 320. A tag directory (TD) 324 may represent data in caches in domain 320.

5 Similarly, processors 306 and 308, and their cores, un-cores, and caches may be associated with domain 330. A TD 332 may represent data in caches in domain 330. A TD 340 may represent data in caches in multiple domains 320 and 330. In other words, the TD structure may be a hierarchy, where TD 324 and TD 332 are on one level of the hierarchy and TD 340 is on the next level. Although only two levels in a TD hierarchy
10 have been illustrated in FIG. 3, other embodiments may include any number of levels in a TD hierarchy.

[0035] If a request for a cache line misses the L1D cache 316, the request may check for the same cache line in the L2 cache 322. If the cache line is not in the L2 cache 322, then the request may continue to check the TD 324 to find out whether the
15 cache line is located in one of the caches controlled by the neighboring cores in the same domain 320 (i.e., the caches in 304). Even if a copy of the cache line is found in a neighboring cache in the same domain 320, there may be other copies of the cache line in other domains (for example, domain 330), which must be accounted for from a cache coherence perspective. Therefore, the request may need to continue to the TD 340, and
20 check if any other domains also have a copy of the cache line.

[0036] With a hierarchical tag directory structure as shown in FIG. 3, it may be advantageous (with regards to reducing energy and latency) to resolve requests for cache lines by only accessing the local domain (for example, 324) without accessing other domains (for example, 332 and 340). To achieve this, a new field per tag entry
25 may be used. In an embodiment, as illustrated in FIG. 4, TDs 424 and 425 have a new field per tag entry (which in turn is associated with a cache line). The new field may be called domain state (475, 477), and may indicate information about a cache line in a manner that minimizes the number of TDs which need to be accessed by a request for a cache line.

30 **[0037]** In an embodiment, the domain state field may include 4 values. The first value, which may be denoted by EJLA, may indicate that the current domain (i.e., the

domain in which the domain state field exists) has the only copies of the cache line in the hierarchy. Referring to FIG. 4, if one of the tag entries in TD 424 had a domain state value of E_LA, that domain state value may indicate that domain 420 has the only copies of the cache line associated with that domain state field. The second value, which may be denoted by SJLA, may indicate that the current domain is the last accessor of the cache line and at least one other domain has or had a copy of the cache line. Referring to FIG. 4, if one of the tag entries in TD 424 had a domain state value of S_LA, that domain state value may indicate that domain 420 is the last accessor of the cache line associated with that domain state field, and at least one other domain (for example, 430) has or had a copy of the cache line. The third value, which may be denoted by S_NLA_S, may indicate that the current domain is not the last accessor of the cache line and all domains are up to date with memory for the cache line associated with that domain state field. Referring to FIG. 4, if one of the tag entries in TD 424 had a domain state value of S_NLA_S, that domain state value may indicate that domain 420 is not the last accessor of the cache line associated with that domain state field, and all domains (420 and 430) are up to date with memory for the cache line associated with that domain state field. The fourth value, which may be denoted by S_NLA_O, may indicate that the current domain is not the last accessor of the cache line and that the global last accessor domain has a copy that is not up to date with memory. Referring to FIG. 4, if one of the tag entries in TD 424 had a domain state value of S_NLA_O, that domain state value may indicate that domain 420 is not the last accessor of the cache line associated with that domain state field, and the global last accessor domain (for example, 430), has a copy of the cache line that is not up to date with memory.

[0038] FIG. 5 illustrates, in an embodiment, various conditions which set the values of the domain state field. In an embodiment, a domain state field's value may be set to E_LA 510 when a read (Rd) request, or request for ownership (RFO), i.e., an attempt to obtain a writable cache line, is issued from the domain state field's domain, resulting in the acquisition of a first copy of the cache line. In an embodiment, a domain state field's value may be set to E_LA 510 when an RFO request is issued from the domain state field's domain, and the cache line is obtained from another domain. In an

embodiment, a domain state field's value may be set to S_LA 520 when a read request is issued from the domain state field's domain, and the cache line is obtained from another domain. In an embodiment, a domain state field's value may be set to S_NLA_S 530 when a read probe from another domain is issued and the cache line in this domain is the same as memory. In an embodiment, a domain state field's value may be set to S_NLA_O 540 when a read probe from another domain is issued and another domain could have a newer version of the cache line than memory. A copy of the modified cache line may then be evicted from the domain state field's domain, and other copies of the cache line may remain in the domain. In an embodiment, a domain state field's value may be set to S_NLA_O 540 when a modified cache line is evicted from the domain state field's domain, at least another copy of the cache line remains in the domain, and the current state of the domain state field is either E_LA or S_LA.

[0039] In an embodiment, the domain state field may include 3 values. The first value, which may be denoted by VI, may indicate that all write and read type requests for the requested cache line may be serviced in the domain state field's domain only. Referring to FIG. 4, if one of the tag entries in TD 424 had a domain state value of VI, that domain state value may indicate that all read and write requests may be fulfilled in domain 420 only. The second value, which may be denoted by V2, may indicate that all read type requests for the cache line may be serviced in the domain state field's domain, but the write requests may not be serviced in the domain state field's domain. Instead, the write requests may be fulfilled by the next level of the tag directory hierarchy. Referring to FIG. 4, if one of the tag entries in TD 424 had a domain state value of V2, that domain state value may indicate that all read requests for its associated cache line may be fulfilled in domain 420, but write requests may have to go to TD 440. The third value, which may be denoted by V3, may indicate that no read type requests nor write request for the cache line may be serviced in the domain state field's domain. Instead, the read/write requests may be fulfilled by the next level of the tag directory hierarchy. Referring to FIG. 4, if one of the tag entries in TD 424 had a domain state value of V3, that domain state value may indicate that all read/write requests for its associated cache line may not be fulfilled in domain 420, but the requests may have to go to TD 440.

[0040] FIG. 6 illustrates, in an embodiment, various conditions which set the values of the domain state field. In an embodiment, a domain state field's value may be set to VI 610 when a read (Rd) request, or request for ownership (RFO), i.e., an attempt to obtain a writable cache line, is issued from the domain state field's domain, resulting in the acquisition of a first copy of the cache line. In an embodiment, a domain state field's value may be set to VI 610 when an RFO request is issued from the domain state field's domain, and the cache line is obtained from another domain. In an embodiment, a domain state field's value may be set to V2 620 when a read request is issued from the domain state field's domain, and the cache line is obtained from another domain. In an embodiment, a domain state field's value may be set to V2 620 when a read probe from another domain is issued and the cache line in domain state field's domain is the same as memory. In an embodiment, a domain state field's value may be set to V3 630 when a read probe from another domain is issued and another domain could have a newer version of the cache line than memory. A copy of the modified cache line may be evicted from the domain state field's domain, and other copies of the cache line may remain in the domain. In an embodiment, a domain state field's value may be set to V3 630 if a modified cache line from the domain state field's domain is evicted, at least another copy of the cache line is present in the domain state field's domain, and the current value of the domain state field is VI.

[0041] A value, as used herein, includes any known representation of a number, a state, a logical state, or a binary logical state. Often, the use of logic levels, logic values, or logical values is also referred to as 1s and 0s, which simply represents binary logic states. For example, a 1 refers to a high logic level and 0 refers to a low logic level. In one embodiment, a storage cell, such as a transistor or flash cell, may be capable of holding a single logical value or multiple logical values. However, other representations of values in computer systems have been used. For example the decimal number ten may also be represented as a binary value of 1010 and a hexadecimal letter A. Therefore, a value includes any representation of information capable of being held in a computer system.

[0042] Moreover, states may be represented by values or portions of values. As an example, a first value, such as a logical one, may represent a default or initial state,

while a second value, such as a logical zero, may represent a non-default state. In addition, the terms reset and set, in one embodiment, refer to a default and an updated value or state, respectively. For example, a default value potentially includes a high logical value, i.e. reset, while an updated value potentially includes a low logical value, i.e. set. Note that any combination of values may be utilized to represent any number of states.

[0043] The embodiments of methods, hardware, software, firmware or code set forth above may be implemented via instructions or code stored on a machine-accessible or machine readable medium which are executable by a processing element. A machine-accessible/readable medium includes any mechanism that provides (i.e., stores and/or transmits) information in a form readable by a machine, such as a computer or electronic system. For example, a machine-accessible medium includes random-access memory (RAM), such as static RAM (SRAM) or dynamic RAM (DRAM); ROM; magnetic or optical storage medium; flash memory devices; electrical storage device, optical storage devices, acoustical storage devices or other form of propagated signal (e.g., carrier waves, infrared signals, digital signals) storage device; etc. For example, a machine may access a storage device through receiving a propagated signal, such as a carrier wave, from a medium capable of holding the information to be transmitted on the propagated signal.

[0044] Reference throughout this specification to "one embodiment" or "an embodiment" means that a particular feature, structure, or characteristic described in connection with the embodiment is included in at least one embodiment of the present invention. Thus, the appearances of the phrases "in one embodiment" or "in an embodiment" in various places throughout this specification are not necessarily all referring to the same embodiment. Furthermore, the particular features, structures, or characteristics may be combined in any suitable manner in one or more embodiments.

[0045] In the foregoing specification, a detailed description has been given with reference to specific exemplary embodiments. It will, however, be evident that various modifications and changes may be made thereto without departing from the broader spirit and scope of the invention as set forth in the appended claims. The specification and drawings are, accordingly, to be regarded in an illustrative sense rather than a

restrictive sense. Furthermore, the foregoing use of embodiment and other exemplary language does not necessarily refer to the same embodiment or the same example, but may refer to different and distinct embodiments, as well as potentially the same embodiment.

CLAIMS

What is claimed is:

1. An apparatus comprising:
 - at least one cache belonging to a first domain of a plurality of domains;
 - at least one cache tag directory belonging to a hierarchy of cache tag directories, wherein the at least one cache tag directory includes at least one tag entry associated with at least one cache line from the at least one cache belonging to the first domain, wherein the at least one tag entry includes a domain state field, wherein the domain state field indicates whether a request from a core can be serviced in a domain of the plurality of domains.
2. The apparatus of claim 1, wherein the domain state field includes at least one of:
 - a first value indicating that the first domain has the only copies of a particular cache line,
 - a second value indicating that the first domain is a last accessor of the particular cache line, and at least one other domain has or had a copy of the particular cache line,
 - a third value indicating that the first domain is not the last accessor of the particular cache line, and all domains are up to date with memory, and
 - a fourth value indicating that the first domain is not the last accessor of the particular cache line, and a global last accessor domain has a copy of the particular cache line which is not up to date with memory.
3. The apparatus of claim 1, wherein the domain state field includes at least one of:
 - a first value indicating that all read requests and all write requests for a particular cache line can only be serviced in the first domain,
 - a second value indicating that all read requests for the particular cache line can be serviced in the first domain, that no write requests for the particular cache line can be serviced in the first domain, and that write requests for the particular cache line

must go to a next level of the hierarchy of cache tag directories, and

a third value indicating that no read requests or write requests for the particular cache line can be serviced in the first domain, and that read requests or write requests for the cache line must go to a next level of the hierarchy of cache tag directories.

4. The apparatus of claim 2, wherein the domain state field is set to the first value if a read request or request for ownership for a cache line associated with the domain state field is issued from the first domain and a first copy of the cache line is obtained and stored in the first domain.

5. The apparatus of claim 2, wherein the domain state field is set to the first value if a request for ownership for a cache line associated with the domain state field is issued from the first domain and a copy of the cache line is obtained from another domain.

6. The apparatus of claim 2, wherein the domain state field is set to the second value if a read request for a cache line associated with the domain state field is issued from the first domain and a copy of the cache line is obtained from another domain.

7. The apparatus of claim 2, wherein the domain state field is set to the third value if a read probe is issued to the first domain from another domain for a cache line in the first domain and the cache line in the first domain is the same as an associated cache line in memory.

8. The apparatus of claim 2, wherein the domain state field is set to the fourth value if a read probe is issued to the first domain from another domain for a cache line in the first domain and a another domain could have a newer copy of the cache line than an associated cache line in memory.

9. The apparatus of claim 2, wherein the domain state field is set to the fourth value if a modified cache line from the first domain is evicted, at least one other copy of the cache line is in the first domain, and a current value of the domain state field is

one of the first value and the second value.

10. The apparatus of claim 3, wherein the domain state field is set to the first value if a read request or request for ownership for a cache line associated with the domain state field is issued from the first domain and a first copy of the cache line is obtained and stored in the first domain.

11. The apparatus of claim 3, wherein the domain state field is set to the first value if a request for ownership for a cache line associated with the domain state field is issued from the first domain and a copy of the cache line is obtained from another domain.

12. The apparatus of claim 3, wherein the domain state field is set to the second value if a read request for a cache line associated with the domain state field is issued from the first domain and a copy of the cache line is obtained from another domain.

13. The apparatus of claim 3, wherein the domain state field is set to the second value if a read probe is issued to the first domain from another domain for a cache line in the first domain and the cache line in the first domain is the same as an associated cache line in memory.

14. The apparatus of claim 3, wherein the domain state field is set to the third value if a read probe is issued to the first domain from another domain for a cache line in the first domain and a another domain could have a newer copy of the cache line than an associated cache line in memory.

15. The apparatus of claim 3, wherein the domain state field is set to the third value if a modified cache line from the first domain is evicted, at least another copy of the cache line is present in the first domain, and a current value of the domain state field is the first value.

16. A method comprising:

setting a value to a domain state field of at least one tag entry in at least a cache tag directory, wherein the at least a cache tag directory belongs to a hierarchy of cache tag directories, and the at least one tag entry is associated with at least one cache line from at least one cache belonging to a first domain of a plurality of domains;

reading the value of the domain state field; and

based on the value of the domain state field, determining whether a request from a core can be serviced in a domain of the plurality of domains.

17. The method of claim 16, wherein the domain state field includes at least one of:

a first value indicating that the first domain has the only copies of a particular cache line,

a second value indicating that the first domain is a last accessor of the particular cache line, and at least one other domain has or had a copy of the particular cache line,

a third value indicating that the first domain is not the last accessor of the particular cache line, and all domains in the hierarchy of cache tag directories are up to date with memory, and

a fourth value indicating that the first domain is not the last accessor of the particular cache line, and a global last accessor domain has a copy of the particular cache line which is not up to date with memory.

18. The method of claim 16, wherein the domain state field includes at least one of:

a first value indicating that all read requests and all write requests for a particular cache line can only be serviced in the first domain,

a second value indicating that all read requests for the particular cache line can be serviced in the first domain, that no write requests for the particular cache line can be serviced in the first domain, and that write requests for the particular cache line must go to a next level of the hierarchy of cache tag directories, and

a third value indicating that no read requests or write requests for the particular cache line can be serviced in the first domain, and that read requests or write requests

for the cache line must go to a next level of the hierarchy of cache tag directories.

19. The method of claim 17, wherein the domain state field is set to the first value if a read request or request for ownership for a cache line associated with the domain state field is issued from the first domain and a first copy of the cache line is obtained and stored in the first domain.

20. The method of claim 17, wherein the domain state field is set to the first value if a request for ownership for a cache line associated with the domain state field is issued from the first domain and a copy of the cache line is obtained from another domain.

21. The method of claim 17, wherein the domain state field is set to the second value if a read request for a cache line associated with the domain state field is issued from the first domain and a copy of the cache line is obtained from another domain.

22. The method of claim 17, wherein the domain state field is set to the third value if a read probe is issued to the first domain from another domain for a cache line in the first domain and the cache line in the first domain is the same as an associated cache line in memory.

23. The method of claim 17, wherein the domain state field is set to the fourth value if a read probe is issued to the first domain from another domain for a cache line in the first domain and a another domain could have a newer copy of the cache line than an associated cache line in memory.

24. The method of claim 17, wherein the domain state field is set to the fourth value if a modified cache line from the first domain is evicted, at least one other copy of the cache line is in the first domain, and a current value of the domain state field is one of the first value and the second value.

25. The method of claim 18, wherein the domain state field is set to the first value

if a read request or request for ownership for a cache line associated with the domain state field is issued from the first domain and a first copy of the cache line is obtained and stored in the first domain.

26. The method of claim 18, wherein the domain state field is set to the first value if a request for ownership for a cache line associated with the domain state field is issued from the first domain and a copy of the cache line is obtained from another domain.

27. The method of claim 18, wherein the domain state field is set to the second value if a read request for a cache line associated with the domain state field is issued from the first domain and a copy of the cache line is obtained from another domain.

28. The method of claim 18, wherein the domain state field is set to the second value if a read probe is issued to the first domain from another domain for a cache line in the first domain and the cache line in the first domain is the same as an associated cache line in memory.

29. The method of claim 18, wherein the domain state field is set to the third value if a read probe is issued to the first domain from another domain for a cache line in the first domain and a another domain could have a newer copy of the cache line than an associated cache line in memory.

30. The method of claim 18, wherein the domain state field is set to the third value if a modified cache line from the first domain is evicted, at least another copy of the cache line is present in the first domain, and a current value of the domain state field is the first value.

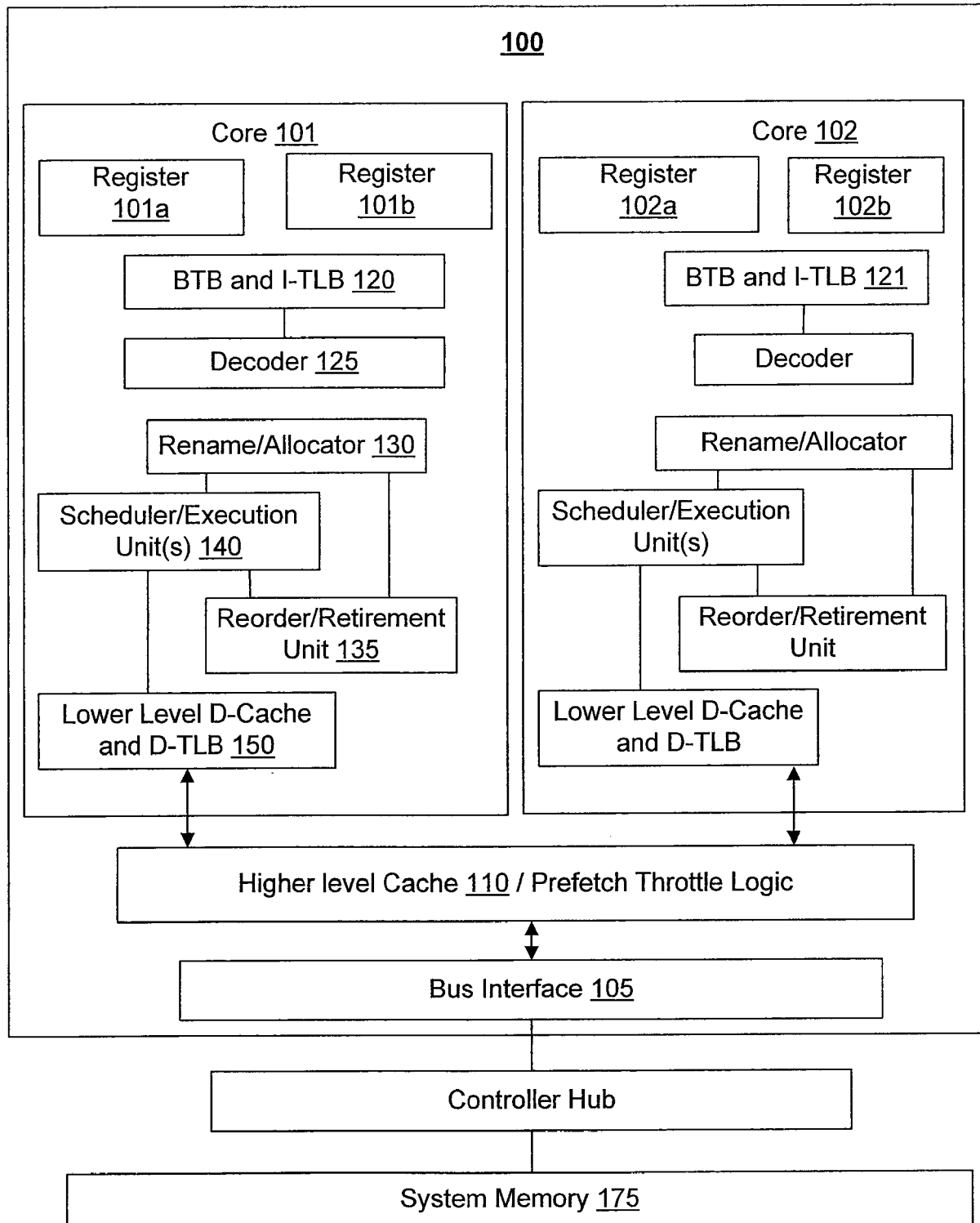


FIG. 1

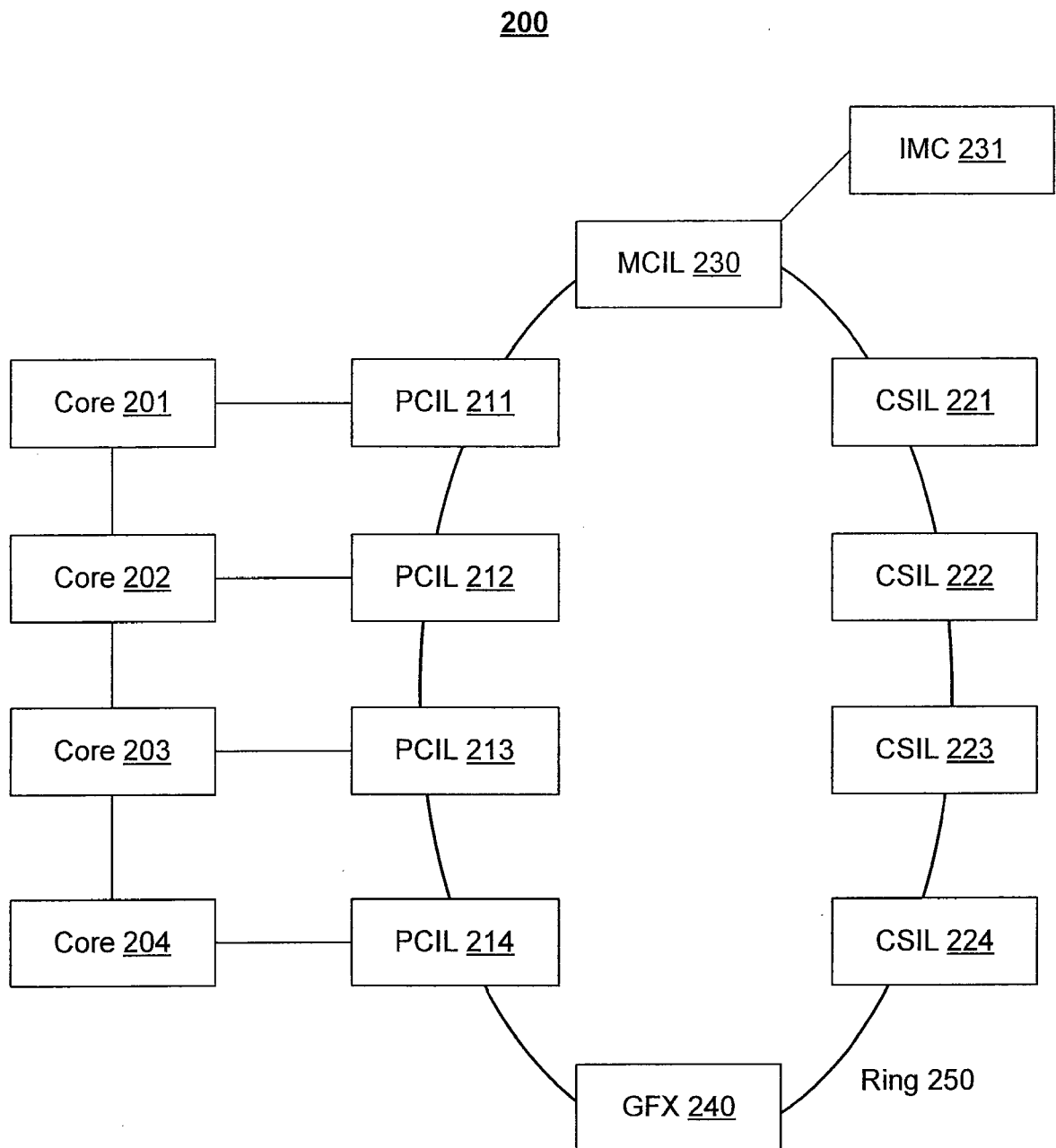


FIG. 2

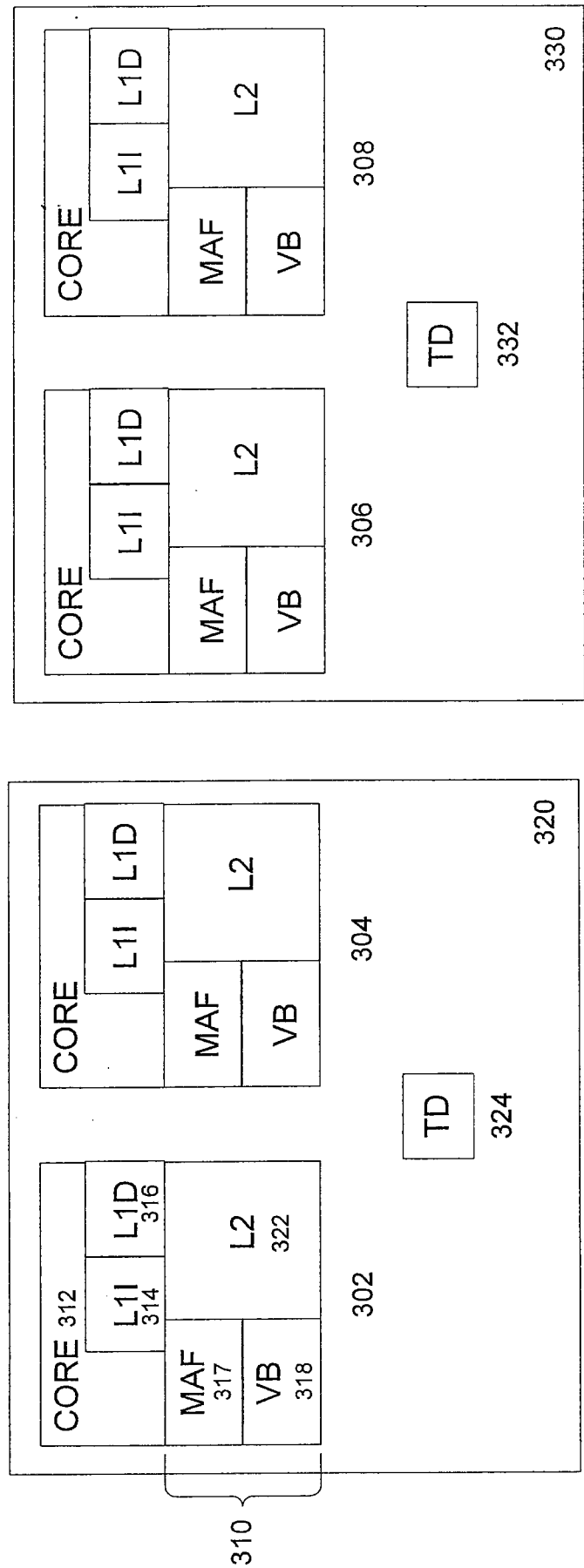


FIG. 3

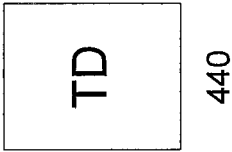
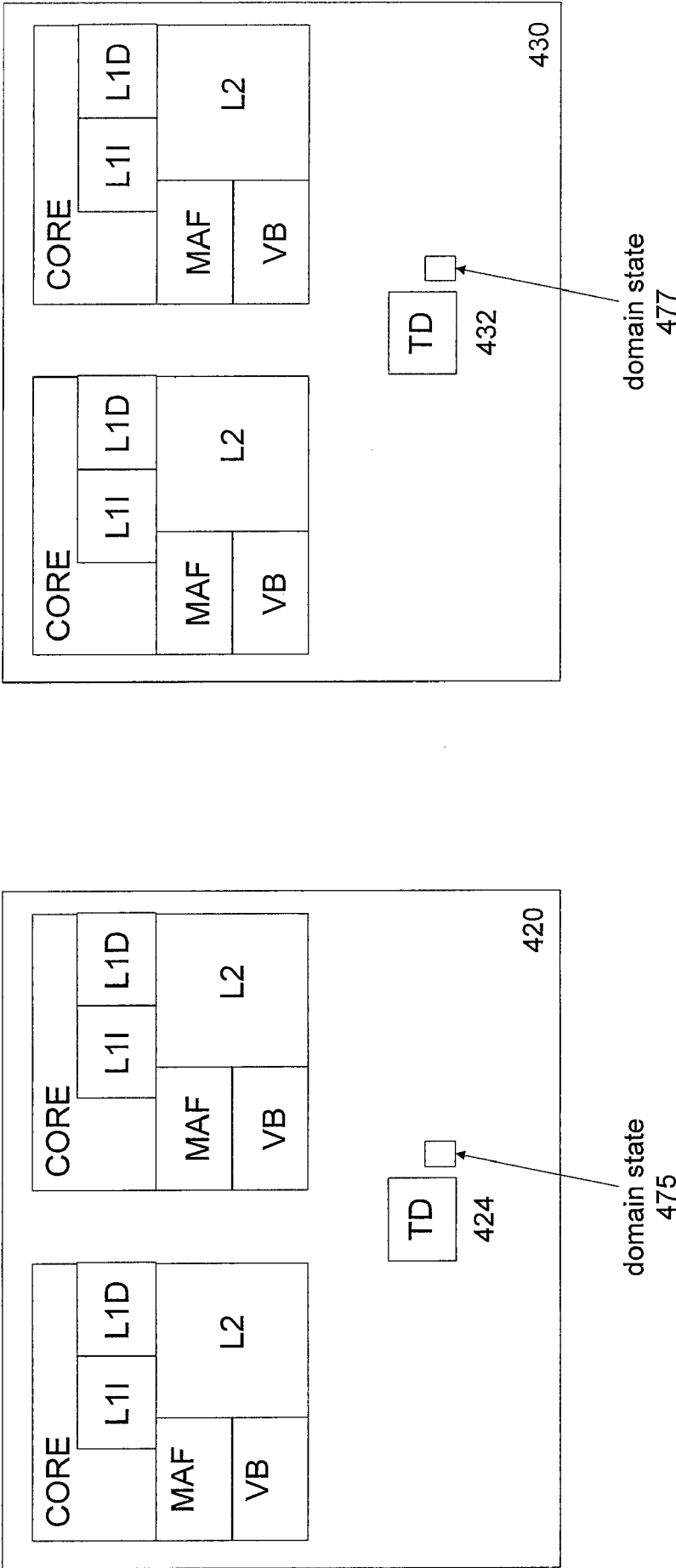


FIG. 4

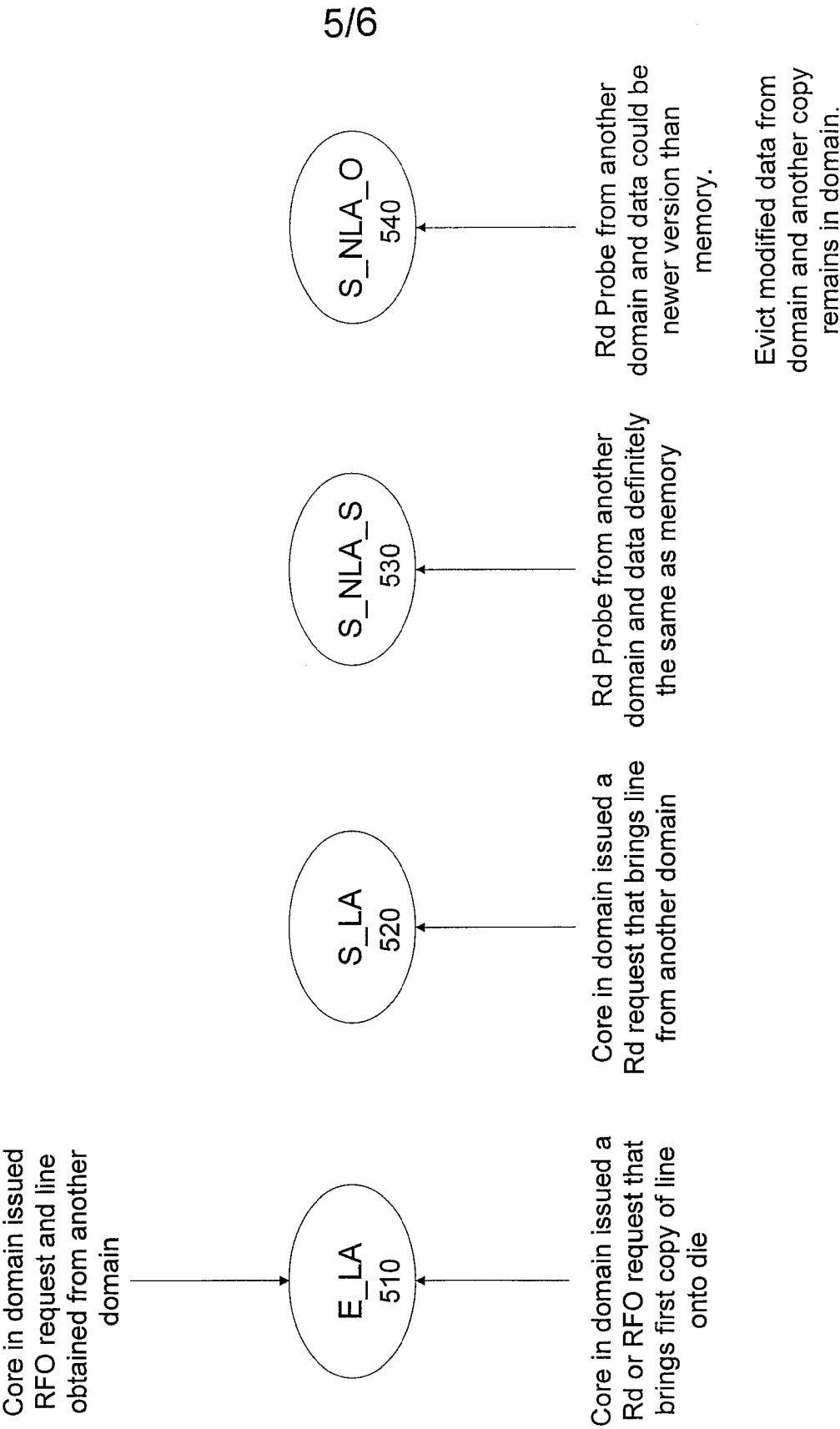


FIG. 5

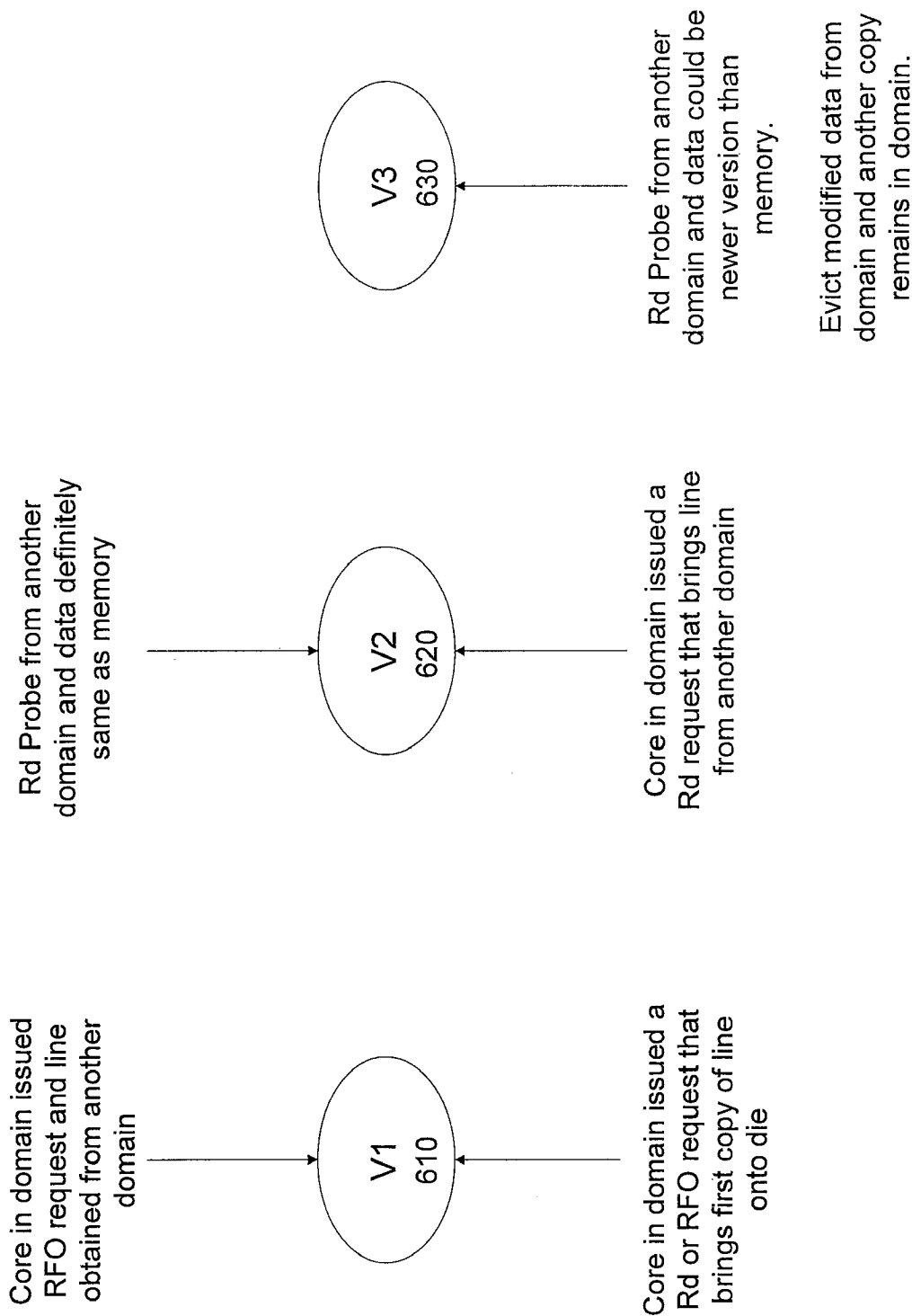


FIG. 6

INTERNATIONAL SEARCH REPORT

International application No.
PCT/US 11/67845

A. CLASSIFICATION OF SUBJECT MATTER
IPC(8) - G06F 13/00 (2012.01)
USPC - 711/141

According to International Patent Classification (IPC) or to both national classification and IPC

B. FIELDS SEARCHED

Minimum documentation searched (classification system followed by classification symbols)
USPC: 711/141

Documentation searched other than minimum documentation to the extent that such documents are included in the fields searched
USPC: 711/141; 711/143; 711/E12.024; 711/E12.043; 711/147; 711/154 (keyword limited - see search terms below)

Electronic data base consulted during the international search (name of data base and, where practicable, search terms used)
PubWEST (PGPB, USPT, USOC, EPAB, JPAB); GOOGLE; Google Scholar
Terms: processors, multiple, cores, cache, domain, coherent, sequential, state, access, local, global, tag, read, write, last, copy.

C. DOCUMENTS CONSIDERED TO BE RELEVANT

Category*	Citation of document, with indication, where appropriate, of the relevant passages	Relevant to claim No.
Y	US 2004/0268059 A1 (Landin et al.) 30 December 2004 (30.12.2004), entire document, especially abstract, para [0003], [0006], [0007], [0076], [0082], [0088], [0092], [0100], [0109], [0199], [0227], [0235], [0283].	1-30
Y	US 2006/0224833 A1 (Guthrie et al.) 05 October 2006 (05.10.2006), entire document, especially abstract, para [0007], [0008], [0044], [0051], [0056], [0059].	1-30
A	US 2004/0010610 A1 (Cypher) 15 January 2004 (15.01.2004), entire document, especially abstract, para [0005], [0006], [0042], [0046], [0050], [0061], [0075].	1-30

☐ Further documents are listed in the continuation of Box C. ☐

* Special categories of cited documents:	"T" later document published after the international filing date or priority date and not in conflict with the application but cited to understand the principle or theory underlying the invention
"A" document defining the general state of the art which is not considered to be of particular relevance	"X" document of particular relevance; the claimed invention cannot be considered novel or cannot be considered to involve an inventive step when the document is taken alone
"E" earlier application or patent but published on or after the international filing date	"Y" document of particular relevance; the claimed invention cannot be considered to involve an inventive step when the document is combined with one or more other such documents, such combination being obvious to a person skilled in the art
"L" document which may throw doubts on priority claim(s) or which is cited to establish the publication date of another citation or other special reason (as specified)	"&" document member of the same patent family
"O" document referring to an oral disclosure, use, exhibition or other means	
"P" document published prior to the international filing date but later than the priority date claimed	

Date of the actual completion of the international search
11 April 2012 (11.04.2012)

Date of mailing of the international search report

27 APR 2012

Name and mailing address of the ISA/US
Mail Stop PCT, Attn: ISA/US, Commissioner for Patents
P.O. Box 1450, Alexandria, Virginia 22313-1450
Facsimile No. 571-273-3201

Authorized officer:
Lee W. Young

PCT Helpdesk: 571-272-4300
PCT OSP: 571-272-7774